

Short Communication

Integrated Information Coefficient Estimated from Neuronal Activity in Hippocampus-Amygdala Complex of Rats as a Measure of Learning Success

Ivan A. Nazhestkin^{1,*,†}, Olga E. Svarnik^{1,2,†}

Submitted: 14 April 2022 Revised: 28 April 2022 Accepted: 5 May 2022 Published: 21 July 2022

Abstract

Background: The goal of the brain is to provide right on time a suitable earlier-acquired model for the future behavior. How a complex structure of neuronal activity underlying a suitable model is selected or fixated is not well understood. Here we propose the integrated information Φ as a possible metric for such complexity of neuronal groups. It quantifies the degree of information integration between different parts of the brain and is lowered when there is a lack of connectivity between different subsets in a system. Methods: We calculated integrated information coefficient (Φ) for activity of hippocampal and amygdala neurons in rats during acquisition of two tasks: spatial task followed by spatial aversive task. An Autoregressive Φ algorithm was used for time-series spike data. Results: We showed that integrated information coefficient Φ is positively correlated with a metric of learning success (a relative number of rewards). Φ for hippocampal neurons was positively correlated with Φ for amygdalar neurons during the learning requiring the cooperative work of hippocampus and amygdala. Conclusions: This result suggests that integrated information coefficient Φ may be used as a prediction tool for the suitable level of complexity of neuronal activity and the future success in learning and adaptation and a tool for estimation of interactions between different brain regions during learning.

Keywords: learning; hippocampus; amygdala; integrated information theory; reconsolidation

1. Introduction

A currently discussed hypothesis is that the brain functions as a past experience-based predictor of future actions in the environment ([1–5], and this idea has a long history [6–11]). This predictive function is possible due to adapted activity of a huge number of neurons.

It is well known that the brain neurons have complex connectivity of two types: some neurons are mostly locally interconnected, some have additionally wide-spread connections — the principle known as "small-world" organization ([12]). The latter ones belong to so called "Rich club" forming a global hub, which interconnects diverse parts of the whole brain [13,14]. Such organization is not limited by morphological connections between neurons but also revealed in functional connectivity. Functional connectivity related to the ability to integrate different parts of the brain into the whole is often analyzed in terms of the information theory [15,16] and, in particular, by the integrated information theory (IIT) proposed by Tononi and his colleagues [17-22], which quantifies an advantage of a cooperative work of brain parts. Φ might be considered as a metric capturing an interior view of a system ("what the system is") instead of describing external characteristics ("what the system does") [23]. IIT was successfully used

in many non-biological complicated systems as a metric of a success [24,25], but was not used to analyze dynamics of neuronal activity in a brain during the period of learning. A resembling approach for functional magnetic resonance imaging data in humans showed an increase of interregional activity correlations between brain regions in contrast with correlations inside regions during different task performance [26].

Integrated information coefficient Φ is calculated using a definition given by authors [19]:

$$\Phi = \sum_{k=1}^{r} H(M_{t,k} \mid M_{t+\Delta t,k}) - H(X_t \mid X_{t+\Delta t}) \quad (1)$$

where H(A|B) is the conditional entropy of variable A with given knowledge of variable B, X_0 and X_1 are the system state vectors at time moments t and $t+\Delta t$ (see below); $M_{t,\ k}$ and $M_{t+\Delta t,\ k}$ are the state vectors of system subsets at time moments t and $t+\Delta t$; and t is the number of subsets. The time here is discrete, it is a set of sequential system states really taken at the same time intervals. Partition by subsets, so-called minimum information partition (MIP), is performed in such a way that subsets must be maximally independent, i.e., the information reduction after a system

¹Institute of Nano, Bio, Information, Cognitive and Socio-Humanitarian Sciences and Technologies (INBICST), Moscow Institute of Physics and Technology, 117303 Moscow, Russian Federation

²Institute of Psychology of Russian Academy of Sciences, 129366 Moscow, Russian Federation

^{*}Correspondence: nazhestkin@phystech.edu; gausseuler60@gmail.com (Ivan A. Nazhestkin)

[†]These authors contributed equally. Academic Editor: Imran Khan Niazi

splitting must be minimal [19].

This approach has some limitations in application to brain neurons. First, due to the computation complexity, the coefficient Φ cannot be precisely estimated for large neuron groups, and it is a discussable question does the approximate value reflect the functional connectivity of a brain. When the state vectors are considered, an information may be lost while converting a complex dynamic of spikes and field potentials into a system state vectors. Finally, this approach does not consider systems with memory, handling an information processing only for a single time step [19,27]. Thus, it is required to check whether the coefficient Φ describe the behavior of a brain.

In this work we used information integration coefficient (Φ) to quantify changes in neuronal activity during learning. Φ was estimated based on two sets of neuronal activity data from hippocampus and amygdala in rats acquiring a spatial aversive task in a linear maze. We showed that Φ for the two structures increased as learning progressed.

2. Materials and Methods

2.1 Calculation of Φ

 Φ was calculated with autoregressive Φ algorithm (Φ_{AR}) [28]. This approach was developed for performing calculations on time-series data, because the estimation of entropies in the definition (1) requires to estimate the whole distribution of data, to gather statistics about 2^N states of a system of N neurons. For real time-series data, this distribution is often underestimated and yields to instability of Φ [28]. To build binary time-series data for calculations, a timeline was divided into equal bins. If there was a spike in the bin, this bin has value 1, and in other case it has value 0. A bin size is used as a compromise, in order to keep all patterns of neuron spiking and avoid excessive information. Such binning method was used for calculation of entropies, for example in [29].

This algorithm has a parameter Δt , which means how much of sequential states are considered (for how many steps the information is generated). This is a temporal scale for a system, it shows a degree to which an information in a network predicts a future network state given an earlier state separated in time by a lag Δt [30]. So, Δt was determined experimentally by calculating Φ with different values of Δt and finding such Δt which gives a maximal average value of Φ for all learning days.

This classical definition (1) cannot be used for calculation of Φ . It is based on estimation of a partition minimizing Φ , which can only be found by a brute-force search. A number of possible ways to partition a system of N elements is the Bell number B_N [31]. As a system size N increases, this number grows faster than a factorial. Thus, another calculation method, so-called autoregressive Φ (Φ_{AR}) was

used [28]:

$$\Phi_{\mathrm{AR}} = \frac{\min\limits_{M} \left(\frac{1}{2} \ln \frac{\det(\Sigma(X))}{\det(\Sigma(E^{X}))} - \sum_{k=1}^{2} \frac{1}{2} \ln \frac{\det(\Sigma(M_{k}))}{\det(\Sigma(E^{M_{k}}))}\right)}{\frac{1}{2} \ln \left[\min\limits_{k} \left\{ (2\pi e)^{|M_{k}|} \det\left(\Sigma\left(M_{k}\right)\right) \right\}\right]}$$
(2)

where $\Sigma(X)$ is the covariation matrix of variable X; E^X is the autoregression residuals (errors of regression that predicts value of X at time moment $t + \Delta t$ based on value of X at time moment t; M_k , $k \in \{0, 1\}$ are the two system subsets, and $|M_k|$ is a size of a subset M_k . Subsets are selected in a way to minimize the resulting Φ_{AR} value. Only two subsets are selected in this method, the number of candidate partitions is defined by a Stirling number of the second kind and grows exponentially. But for a large number of neurons a calculation in a considerable time is still impossible. Thus, an approximation was applied. Experimentally we found that the limit of 15 neurons is the point when computations become intractable. To perform calculations, at each session we selected 15 neurons with the most variance of activity, as it was done in [26]. In previous works, other approximations based on simplified partition were used [32].

In some cases, this algorithm could not return a value. It is known [26], that it happens when some neurons had little or no variance in their activities — bins were almost always on or almost always off. In this case, the one neuron with the least variance was excluded from analysis and calculation was performed again, and so on. If an algorithm was unable to calculate Φ_{AR} on any neurons in some period, this period was excluded from statistical analysis. Each obtained Φ_{AR} value was normalized by a number of neurons used in this calculation to make it possible to compare obtained values and investigate its dynamics through each animal learning sessions.

To achieve a higher time resolution, each session was divided into 8 equal periods and information integration coefficient Φ_{AR} was also calculated for each period. Calculations were performed in MATLAB. Neural and behavioral data manipulations were performed with FMAToolbox open-source library (http://fmatoolbox.sourceforge.net) [33].

2.2 Behavioral Paradigm

We used an open dataset [34,35] in order to analyze behavior and neuronal activity during learning, so details of the experimental design may be found in [35]. Animal experiments were approved by the Institutional Animal Care and Use Committee (IACUC) at New York University Medical Center, as it was mentioned in an original manuscript [35]. In short, male Long-Evans rats (N=3, 300 g, 3 months) were initially trained a simple spatial task in a linear track. During behavior sessions the animals ran from one end of the track to the other to get water from a



All animals, hippocampus

0.15 r=0.3792, p < 0.0001, **** 0.00 0.005 0.010 0.015 0.020 Φ_{AR}

All animals, amygdala

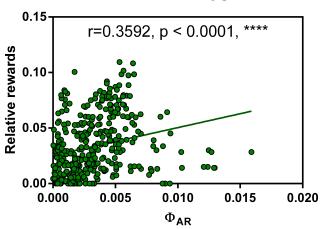


Fig. 1. Integrated information coefficient Φ_{AR} for all animals in hippocampus and amygdala vs. number of rewards in corresponding periods. X axis: Φ_{AR} , Y axis: relative number of rewards, each point represents 1/8 length period in each session. Left: for hippocampal neurons; right: for amygdala neurons. Statistical significance is given taking into account the Bonferroni correction.

water well. After shaping or initial training sessions, microelectrode arrays were implanted into the amygdala and the dorsal hippocampus. After a rehabilitation period of 5 days, animals started learning the second procedure. Two air-puff sites were located at the equal distance from the middle of the track and applied at one of two sites when the rat moved in a particular direction. The direction and the site varied pseudorandomly across training days.

2.3 Data Analysis

Along with information integration coefficient we quantified learning success for every period of each session. Learning success was assessed as a number of rewards in a period divided by a length of a period. Correlations between the two variables in each period were calculated using Spearman's correlation; we also calculated Spearman's ρ using Student's t test. The Bonferroni correction for multiple comparisons was performed.

3. Results

Two animals acquired the complete task and showed learning progress, and the third did not exhibit significant progress and its performance decreased at the beginning of the second half of a learning process. At most each of them performed about 100–150 rewarded trials per day (session). The number of learning days for each rat was different, but enough to estimate the presence or lack of a progress.

An integrated information coefficient was calculated for each day and each period. A bin size was selected as 0.2 seconds (see Materials and method section). The characteristic temporal scales Δt for all animals and the brain regions were determined (Table 1).

We found a positive correlation between the integrated information coefficient Φ_{AR} for the neural activity in the brain areas and the number of rewards obtained by ani-

Table 1. Number of steps for information generation used to calculate an integration information coefficient Φ_{AR} for each animal and brain area

animai and brain area.			
Brain regions	Animal 1	Animal 2	Animal 3
Hippocampus	$\Delta t = 8$	$\Delta t = 10$	$\Delta t = 11$
Amygdala	$\Delta t = 8$	$\Delta t = 12$	$\Delta t = 9$

mals. Such positive correlations were evident in both hippocampus (r = 0.3792, p = 0.0001, ***), and amygdala (r = 0.3592, p = 0.0001, ***) (Fig. 1).

We reasoned that those behavioral adaptations were the function of the whole brain activity and estimated mutual relationships between the two brain areas calculating correlations between integrated information coefficient Φ_{AR} for neuronal activity in the hippocampus and integrated information coefficient Φ_{AR} for neuronal activity in the amygdala. Φ_{AR} in the hippocampus and the amygdala were significantly correlated for two animals exhibited a steady learning progress: Spearman r=0.4299, p=0.0001 for the first animal, and Spearman r=0.4206, p=0.0001 for the second animal (Fig. 2).

We also found that Φ_{AR} for the hippocampal neurons correlated with Φ_{AR} for the amygdala neurons the day the maximal number of rewards occurred (maximal learning progression) and the day before but not the day after it (r=0.7678, p<0.0001, ****; r=0.5520, p<0.0001, ****; <math>r=0.1343, p=0.2608, ns; correspondingly) (Fig. 3). Maximal learning progression occurred on the 10-th day for the first rat, on the 10-th day for the second rat and on the 12th day for the third rat. We also figure out a correlation between Φ_{AR} and number of rewards on the day before and after this point, separately. Φ_{AR} for the amygdala and the hippocampus neurons were significantly correlated with a relative number of rewards only before the maximal pro-



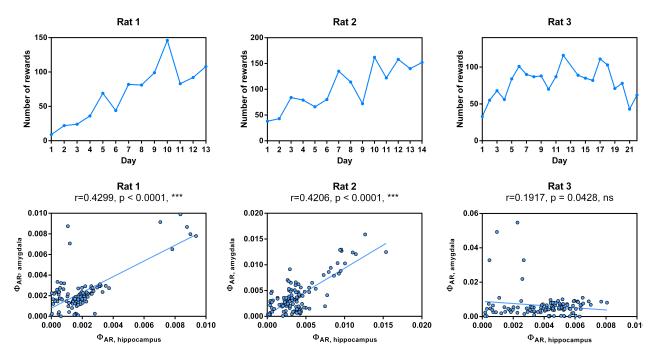


Fig. 2. Learning progress and integrated information coefficient Φ_{AR} for each animal in hippocampus and amygdala. Top: learning progress of animals. X axis: day of learning, Y axis: number of rewards. Bottom: Correlation of Φ_{AR} values for hippocampus and amygdala in the periods. X axis: Φ_{AR} for neural activity in hippocampus, Y axis: Φ_{AR} for neural activity in amygdala, each point represents 1/8 length period of each session. Statistical levels are calculated taking into account the Bonferroni correction for multiple comparisons.

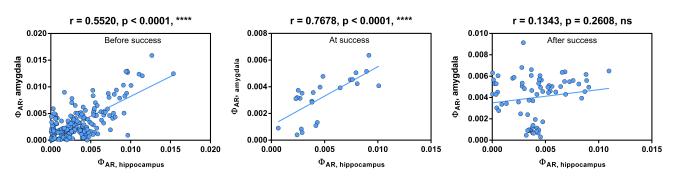


Fig. 3. Information coefficient Φ_{AR} for all animals in hippocampus and amygdala before a day of maximal learning progression. Top row: days before a day of maximal learning progression and at such day; bottom row: days after a day of maximal learning progression. X axis: Φ_{AR} for neural activity in hippocampus, Y axis: Φ_{AR} for neural activity in amygdala, each point represents 1/8 length period of each session. Statistical levels are calculated taking into account the Bonferroni correction for multiple comparisons. Statistical significance was found only for days before a day of maximal learning progression and at this day.

gression day, but not after it. For the hippocampus, the values were Spearman r = 0.4268, p = 0.0001 (before) and Spearman r = 0.09521, p = 0.4263 (after). For the amygdala the values were Spearman r = 0.3933, p = 0.0001 (before) and Spearman r = 0.2372, p = 0.0118 (after).

4. Discussion

Our results suggest that the concept of integrated information, as formalized by the integrated information coefficient Φ , can be usefully applied to assess learning progress. Unlike conventional assessment tools based on

observable values with limitations of being environment-specific and task-specific [24] the integrated information measure might be used as a predictor of near future success in learning or as an indicator of the optimal brain functioning during adaptation. Getting to the point of desired outcome by a particular neuronal pattern is selected and fixated. Given that the integrated information metric was originally proposed for consciousness level assessments [21,36,37], these results open a question what is assessed by this measure.

The metric of integrated information may also be used for establishing interactions among various brain regions



during learning. Similar estimations were quantified as an "information flow" earlier [15,16,38]. In this study, we showed a degree of joint regional contribution underlying behavioral adaptations to the environmental demands. Behavioral paradigm applied in this study consists of two parts: simple spatial task in the linear track mostly based on hippocampal place neurons and spatial aversive task presumably subserved by the amygdala neurons. Our results suggest that neurons of both regions participate in the integration needed for successful learning of the second skill. These results are also in accordance with reconsolidation research findings that indicated that any learning is based on the earlier acquired experience which is required to reconsolidate again [39-41]. The integrated information approach opens a new way for investigation of reconsolidation processes and suggests a possibility to find if reconsolidation processes are regionally restricted to any brain zones under any given learning.

Another possibility is to assess functional connectivity using this approach. As we showed correlated complexity of activity varied through learning progress and dropped on the day after the maximal progression, which might be related to consolidation/reconsolidation during sleep [34,42].

In this study we proposed a new approximation of Φ calculation. As mentioned in Materials and Methods section, we used an approximation preserving the algorithm of selection of a minimum information partition, but sacrificing the coverage of all neurons. We supposed that if neurons have a multi-level organization [43], then important structural changes can be detected even in small groups of neurons, and these changes can approximately reflect changes of a whole anatomical region. This approximation successfully demonstrated a correlation with a classical success metric in 48 days of learning. Thus, a 15neuron approximation is acceptable for performing calculations of Φ . More investigations are required to compare Φ calculated with the proposed approximation and Φ calculated with known approximations. It is necessary to note that given technical advances for large neuronal population registration, it will be a challenging task to track the evolution of Φ of a bigger number of neurons in the same way.

Author Contributions

OS conceived and designed the experiments; IN performed calculations and analyzed the data; IN and OS wrote the paper.

Ethics Approval and Consent to Participate

For this paper, an open dataset was used. Animal experiments were conducted in Gyorgy Buzsaki lab and approved by the Institutional Animal Care and Use Committee (IACUC) at New York University Medical Center.

Acknowledgment

Funding support from RFBR (Project No 20-34-90023) is greatly appreciated.

Funding

This study was supported by grant from the Russian foundation for basic research (Project No 20-34-90023).

Conflict of Interest

The authors declare no conflict of interest.

References

- [1] Barrett LF, Simmons WK. Interoceptive predictions in the brain. Nature Reviews Neuroscience. 2015; 16: 419–429.
- [2] Yuste R, MacLean JN, Smith J, Lansner A. The cortex as a central pattern generator. Nature Reviews Neuroscience. 2005; 6: 477–483.
- [3] Stephan KE, Harrison LM, Kiebel SJ, David O, Penny WD, Friston KJ. Dynamic causal models of neural system dynamics: current state and future extensions. Journal of Biosciences. 2007; 32: 129–144.
- [4] Clark A. Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behavioral and Brain Sciences. 2013; 36: 181–204.
- [5] Buzsáki G. The brain from inside out. Oxford University Press: Oxford. 2019.
- [6] Merleau-Ponty M. Phenomenology of perception (C. Smith, Trans.). Routledge Classics: New York. 1945.
- [7] Merleau-Ponty M. The Cambridge Companion to Merleau-Ponty. Cambridge University Press: Cambridge. 2005.
- [8] MacKay DM. Ways of looking at perception. Models for the perception of speech and visual form (pp. 25–43). MIT Press: Cambridge. 1967.
- [9] Anokhin PK. The problem of decision-making in psychology and physiology. Voprosy Psychologii. 1974; 4: 21–29.
- [10] Shvyrkov VB. Behavioral specialization of neurons and the system-selection hypothesis of learning (pp. 599–611). Human Memory and Cognitive Capabilities: Amsterdam. 1986.
- [11] Alexandrov YI, Grechenko TN, Gavrilov VV, Gorkin AG, Shevchenko DG, Grinchenko YV, et al. Formation and realization of individual experience. Neuroscience and Behavioral Physiology. 1997; 27: 441–454.
- [12] Liao X, Vasilakos AV, He Y. Small-world human brain networks: Perspectives and challenges. Neuroscience & Biobehavioral Reviews. 2017; 77: 286–300.
- [13] van den Heuvel MP, Sporns O. Rich-Club Organization of the Human Connectome. Journal of Neuroscience. 2011; 31: 15775–15786.
- [14] Petersen S, Sporns O. Brain Networks and Cognitive Architectures. Neuron. 2015; 88: 207–219.
- [15] Kirst C, Timme M, Battaglia D. Dynamic information routing in complex networks. Nature Communications. 2016; 7: 11061.
- [16] Palmigiano A, Geisel T, Wolf F, Battaglia D. Flexible information routing by transient synchrony. Nature Neuroscience. 2017; 20: 1014–1022.
- [17] Tononi G. Complexity and coherency: integrating information in the brain. Trends in Cognitive Sciences. 1998; 2: 474–484.
- [18] Tononi G. An information integration theory of consciousness. BMC Neuroscience. 2004; 5: 1–22.
- [19] Balduzzi D, Tononi G. Integrated information in discrete dynamical systems: motivation and theoretical framework. PLOS Computational Biology. 2008; 4: e1000091.
- [20] Tononi G. Integrated information theory of consciousness: an



- updated account. Archives Italiennes de Biologie. 2012; 150: 56-90.
- [21] Casali AG, Gosseries O, Rosanova M, Boly M, Sarasso S, Casali KR, et al. A Theoretically Based Index of Consciousness Independent of Sensory Processing and Behavior. Science Translational Medicine. 2013; 5: 198ra105.
- [22] Tononi G, Boly M, Massimini M, Koch C. Integrated information theory: from consciousness to its physical substrate. Nature Reviews Neuroscience. 2016; 17: 450–461.
- [23] Niizato T, Sakamoto K, Mototake YI, Murakami H, Tomaru T, Hoshika T, et al. Finding continuity and discontinuity in fish schools via integrated information theory. PLoS ONE. 2020; 15: e0229573.
- [24] Edlund JA, Chaumont N, Hintze A, Koch C, Tononi G, Adami C. Integrated information increases with fitness in the evolution of animats. PLOS Computational Biology. 2011; 7: e1002236.
- [25] Engel D, Malone TW. Integrated information as a metric for group interaction. PLoS ONE. 2018; 13: e0205335.
- [26] Shine J, Bissett P, Bell P, Koyejo O, Balsters J, Gorgolewski K, et al. The Dynamics of Functional Brain Networks: Integrated Network States during Cognitive Task Performance. Neuron. 2016; 92: 544–554.
- [27] Barrett AB, Mediano PA. The Phi measure of integrated information is not well-defined for general physical systems. Journal of Consciousness Studies. 2019; 26: 11–20.
- [28] Barrett AB, Seth AK. Practical measures of integrated information for time-series data. PLOS Computational Biology. 2011; 7: e1001052.
- [29] Strong SP, Koberle R, de Ruyter van Steveninck RR, Bialek W. Entropy and Information in Neural Spike Trains. Physical Review Letters. 1998; 80: 197–200.
- [30] Isler JR, Stark RI, Grieve PG, Welch MG, Myers MM. Integrated information in the EEG of preterm infants increases with family nurture intervention, age, and conscious state. PLoS ONE. 2018; 13: e0206237.
- [31] Rota G. The Number of Partitions of a Set. The American Math-

- ematical Monthly. 1964; 71: 498-504.
- [32] Toker D, Sommer FT. Information integration in large brain networks. PLoS Computational Biology. 2019; 15: e1006807.
- [33] Zugaro M, Todorova R, Girardeau G, Cei A, El Kanbi K. FMA-Toolbox. Available at: http://fmatoolbox.sourceforge.net (Accessed: 17 April 2021).
- [34] Girardeau G, Inema I, Buzsáki G. Simultaneous large-scale recordings in dorsal hippocampus, basolateral amygdala and neighbouring deep nuclei and structures in rats performing a spatial aversive task and sleeping. 2017. Available at: http://dx.doi.org/10.6080/K0MS3QXD (Accessed: 1 May 2022).
- [35] Girardeau G, Inema I, Buzsáki G. Reactivations of emotional memory in the hippocampus-amygdala system during sleep. Nature Neuroscience. 2017; 20: 1634–1642.
- [36] Alkire MT, Hudetz AG, Tononi G. Consciousness and Anesthesia. Science. 2008; 322: 876–880.
- [37] King J, Sitt J, Faugeras F, Rohaut B, El Karoui I, Cohen L, et al. Information Sharing in the Brain Indexes Consciousness in Noncommunicative Patients. Current Biology. 2013; 23: 1914–1919
- [38] Avena-Koenigsberger A, Misic B, Sporns O. Communication dynamics in complex brain networks. Nature Reviews Neuroscience. 2018; 19: 17–33.
- [39] Dudai Y. The Restless Engram: Consolidations never End. Annual Review of Neuroscience. 2012; 35: 227–247.
- [40] Alberini CM, LeDoux JE. Memory reconsolidation. Current Biology. 2013; 23: R746–R750.
- [41] Svarnik OE, Anokhin KV, Aleksandrov YI. Experience of a first, "Whisker-Dependent," Skill Affects the Induction of c-Fos Expression in Somatosensory Cortex Barrel Field Neurons in Rats on Training to a second Skill. Neuroscience and Behavioral Physiology. 2015; 45: 724–727.
- [42] Lewis PA, Durrant SJ. Overlapping memory replay during sleep builds cognitive schemata. Trends in Cognitive Sciences. 2011; 15: 343–351.
- [43] Park H, Friston K. Structural and Functional Brain Networks: from Connections to Cognition. Science. 2013; 342: 1238411.

